1940

A study of the comprehensive examination administered to graduate students in the Department of Education, Massachusetts State College, 1939.

John A. Fitzgerald
University of Massachusetts Amherst

A STUDY OF THE COMPREHENSIVE EXAMINATION ADMINISTERED TO GRADUATE STUDENTS IN THE DEPARTMENT OF EDUCATION MASSACHUSETTS STATE COLLEGE, 1939

FITZGERALD - 1940

A STUDY OF THE COMPREHENSIVE EXAMINATION ADMINISTERED TO GRADUATE STUDENTS IN THE DEPARTMENT OF EDUCATION MASSACHUSETTS STATE COLLEGE, 1939

by

JOHN A. FITZGERALD

THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

MASSACHUSETTS STATE COLLEGE

AMHERST, MASSACHUSETTS

1940

TABLE OF CONTENTS

TABLE OF CONTENTS

INDEX OF TABLES

A STUDY OF THE COMPREHENSIVE EXAMINATION ADMINISTERED TO GRADUATE STUDENTS IN THE DEPARTMENT OF EDUCATION

MASSACHUSETTS STATE COLLEGE, 1939

CHAPTER I

THE INTRODUCTION

CHAPTER I

THE INTRODUCTION

The preparation of students for secondary school teaching is coming to be a very important function of colleges and universities throughout the United States. During the nineteenth century the secondary schools of this country contained a relatively small number of pupils who were more or less homogeneous in family background, in intelligence, and in the type of training desired. Since the type of training was essentially academic in character and the pupils homogeneous in nature, it was deemed sufficient that the teacher be an expert in the subject-field to be taught. With the twentieth century the picture has changed drastically so far as the numbers, the characteristics and the needs of our secondary school pupils are concerned. Now a much wider type of training is deemed necessary for teachers. Just what this training should consist in is a matter of considerable controversy and of diverse practices. Several important national committees are working on this problem. In all of their discussions they immediately and necessarily come upon the problem of how to measure teaching ability. The need for this is shown in two major phases of educational work.

1. The Need for Evaluation of Teaching Ability: The first phase of educational work demanding some measure of teaching ability is concerned with teachers in-service. How can we secure teachers for new positions? How can we

promote teachers according to merit? Which teachers should be requested to undertake improvement courses? When should a teacher be admitted to tenure? Are teachers comparable so that we can undertake the equivalent group method of experimentation? It is obvious that in questions of this kind the answer depends upon some accurate measure of teaching ability. Lacking this measure leads administrators and research men into the use of inferior and inadequate answers to such problems.

2. The Need in Prognosis of Teaching Ability: The second phase of educational work demanding some measure of teaching ability is concerned with the training of new secondary school teachers. How can we best select candidates for teacher training classes? How can we successfully recommend students as being good teachers? How can we evaluate the probable success of our training methods? What courses are most functional in preparing these students for teaching? These, and many other problems are constantly before the teacher training departments, both undergraduate and graduate, and the answer as above depends upon obtaining some measure of teaching ability or success. Given this and the possibilities of improvement in present preparation programs are numerous; lacking it there is little save present trial and error methods to fall back upon. It is true that rational subjective thinking and judgment will aid greatly in setting up criteria and defining values, but in the end the need for a definite measure of success in teaching becomes paramount.

What is involved in the search for such a measure?

3. The Difficulty in Measuring Teaching Success: The list of possible factors involved in teaching success is so varied and the difficulty of defining some of these factors so great that the lack of such a measure, or measures, is not surprising. One person says we need educated men and women. But what is an educated man or woman? How educated? In what field? To what degree? Broadly or intensively? Another says we need men and women with good personalities. But what kind of personality? And how shall be judged its presence? And is there only one kind of personality successful in teaching? And so the questioning continues. It is not to be inferred from all of this that the search for a measure of teaching success is doomed to failure but that the search is an extremely difficult one, and much time still must be given to it. In the meantime the training departments have been concentrating upon one phase of possible teaching success which seems to offer the greatest present possibility of value. This measure is that of success in academic and professional courses on the college level.

4. The Comprehensive Examination: A few years ago our sole measure of academic or professional success was the marks which students obtained in their college courses. With the elective system came the situation that students were taking different series and schedules of courses in preparing for their degree and the possibility that marks in various courses were not comparable. To meet this situation came the comprehensive examination designed to measure the students in

those abilities deemed important and valid in anyone preparing in a given field or for a given profession. Of necessity these examinations were academic in character, but they were an improvement over the average class grade technique in that they were a cross-section of the abilities required in all specific subjects and thus the student was permitted to choose the courses best suited to his individual needs. Many subject-fields have taken over the comprehensive examination idea with considerable success, and of late years it is appearing in teacher training departments. The first one for teachers at Massachusetts State College was given to candidates for the master's degree with a major in education in 1939. It is the purpose of this study to analyze the results of that examination with a view to its improvement and with the aim of discovering in so far as possible just what it contributes toward a future measure of teaching success.

CHAPTER II

RELATED LITERATURE

CHAPTER II

RELATED LITERATURE

Literature related to this study is extensive. In order to obtain some organization, the literature will appear under several classifications.

Teacher Education: With regard to Teacher Education Chamberlain (1) states that the modern program of teacher preparation is usually directed toward three general objectives:

> 1. the mastery of the subject matter and skills he is to teach, and an understanding of the ideals, habits, and attitudes he is to inculcate in pupils.
>
> 2. to make the teacher a thorough student of the child, the learning process and teaching procedures, and a loyal member of the profession.
>
> 3. to make the teacher a person of general culture and refinement.

Phillips (2) in his account of the plan at the University of North Carolina states that with the head of the Education Department as chairman, the division of teacher training is composed of members of the Education Department in the college of arts and sciences. This division of teacher training is empowered to create an advisory board, set up standards for the selection of students, approve courses in the teacher-training program, and formulate and administer programs for the

---

1. Chamberlain, L.M. The Teacher and School Organization. New York: Prentice-Hall Inc. 1936. pp. xxviii ± 656. Chapter 5.

2. Phillips, G.B. Teacher Training Deviates at North Carolina. Nation's Schools 22:29, December 6, 1938.

preparation of public school teachers. All students preparing for public school teaching are under their guidance.

Another field to receive considerable attention of late is that of Student Rating and the Qualifying Examination: Major (3) draws the conclusion that with university students a percentile rank of 30 on the Ohio State Psychological Test is representative of sufficient intelligence for successful teaching, or that the test itself is invalid as a test of the kind of intelligence most used by the teacher, or that the present methods of measuring teacher ability are inadequate.

Eldred (4) in a study of three classes of prospective secondary school teachers at Cornell, compared these prospective teachers with their total classes and found: 1) that in intelligence these prospective teachers ranked above their college mates, 2) that the weighted averages of the subjects to be taught are even higher than the general average of the group in all subjects, 3) that the group is above average in high school ratings of personal qualities, 4) that a greater than normal percentage was elected to scholastic honorary societies, 5) that almost every individual was engaged in some extra-class activity, 6) that of 127 students, 64 held 125 scholarships.

---

3. Major, C.L. Percentile Ranking on the Ohio State University Psychological Test as a Factor in Forecasting the Success of Teachers in Training. School and Society, 47:582-4, April 30, 1938.

4. Eldred, Lewis: Traits of a Group of Prospective Teachers in the College of Arts and Sciences, Cornell University. School and Society, 50:477-50, October 7, 1939.

Mathews (5) in his item analysis of measures of teaching ability used as a criterion of teaching success, a composite index derived from measurable changes produced in pupils by teachers of the social studies. He assumes that the best examination measures of teaching ability are those items which, individually, possess a high discriminating power, i.e., items in which there is a tendency for subjects with high criterion scores to have a greater proportion of desirable responses to the items under investigation than do subjects with low criterion scores. His general conclusions are: 1) that the findings of the investigation would seem to cast serious doubt upon the validity of the tests therein studied as measures of teaching ability when pupil change is used as the criterion of teaching ability, 2) that the results offer very little in the way of clues to types or categories of items for the refinement of the tests employed or the creation of new ones, 3) that it is not the opinion of the writer that the measurement of teaching ability is a hopeless task. It would seem, however, that some of the energy being put into the problem might well be directed to a more careful study of existing measures themselves and the development of more refined measures.

Troyer (6) states that, at Syracuse in about 80 percent of the cases of refusal for admission to the selected group

---

5. Mathews, L.H.: An Item Analysis of Measures of Teaching Ability. Journal of Educational Research 33:576-81, April, 1940.

6. Troyer, Maurice E.: The Selection of Students for the Profession of Teaching. Journal of Educational Research, 33:581-94, April, 1940.

of teacher trainees, a constellation of factors operates. Undesirable personality operated alone in 9 percent of the cases and in combination with other factors in 36 percent of the cases. Evaluation on this factor represents the combined judgment of four members of the Enrollment Committee, the Dean of Men, and the Dean of Women, and is based on individual interviews and significant data on extra curricular activities. Scholarship operated in 8 percent of the cases as a single factor and in 58 percent of the cases in combination with other factors. Speech operated alone in two percent of the cases, and in combination with other factors, in three percent of the cases.

In a survey of recent developments in the pre-training selection and guidance of teachers, Steiner and Von Haden (7) mention the conclusions of other authors: Kriner (8) claims, among other things: 1) that no single factor is a reliable index to predicting teacher success, 2) that ability in mathematics, Latin, and the sciences are more reliable for predicting success than is ability in the so-called "arts", 3) that the well-informed, properly educated student with the proper native ability will become a successful teacher.

---

7. Steiner, R.W. and Von Haden, H.I.: The Pretraining Selection and Guidance of Teachers, Journal of Educational Research 33:321-350, January, 1940.

8. Kriner, H.L.: Second Report on a Five Year Study of Teachers' College Admissions. Educational Administration and Supervision, 21:56-60, January, 1935.

Sandiford (9) claims among other things that ability of students in teaching is not closely related to achievement in special subjects, and that neither the comprehensive tests nor the specialist's examinations proved to be of value in predicting teaching ability. He points out that while the correlations are all low, they are also positive, showing that the good student is slightly more apt to succeed as a teacher than the poor student. He also claims that the marks of critic teachers are subject to the variance of individual subjective evaluations. Steiner and Von Haden summarize trends among which are: 1) recognition that no single measure thus far employed has sufficient validity in terms of the criteria used thus far to warrant its acceptance as a valid measure of teaching success, 2) a realization that valid instruments of prediction cannot be established until a valid and reliable criterion of teaching success is determined, 3) acceptance of the principle that selection must be shifted from the basis of subjective opinion to that of scientific investigation, 4) recognition of the value of cumulative records as a basis for selection, 5) scientific investigations which attempt to establish the validity and reliability of the instruments to be used as predictive measures.

Bossing (10) in writing of the testing and selection of

---

9. Sandiford, P., et al: Forecasting Teaching Ability. Elementary School Journal 38:326-9, January, 1938.

10. Bossing, Nelson L. Teacher Aptitude Tests and Teacher Selection. Research in Higher Education, Washington; U.S.Dept. of Interior, Office of Ed. Gen., Printing Office, 1932. pp 117-133.

teachers, quotes Ullman (11) who found a correlation of .30 between professional courses in education and teaching success, and Whitney (12) who found a much higher relationship between teaching success and professional marks than with academic marks. This "much higher relationship" was a correlation of .143 for professional marks. Bossing mentions (in his study) that the criterion of teaching is based upon the judgments of superintendents and principals who supervised for one or more years the work of the teachers judged. Correlations between the criterion and 1) cadet teaching grades, 2) professional education grades, and 3) all academic grades not including grades in professional education courses are:

1) .687 ± .072

2) .188 ± .056

3) .172 ± .088

Of these correlations, the only significant item is that of practice teaching grades. Among his conclusions Bossing states that apparently little confidence can be placed in academic grades, professional grades, or intelligence ratings, as a prediction of teaching success. A corollary conclusion to the above would be that an analysis of the criterion used in this study does not offer much suggestion as to what particular elements should enter into the formation of teach-

---

11. Ullman, Roy R., The Prognostic Value of Certain Factors Related to Teaching Success, Ashland, Ohio. The A.L. Garber Co. 1931. 133 p.

12. Whitney, F.L. The Prediction of Teaching Success. Journal of Educational Research, Monograph #6, Bloomington, Illinois: Public School Publishing Co. 1924. 85 p.

ing aptitude tests. Among his general conclusions, Bossing states that practice teaching alone has shown a significant relationship to later teaching success. Unfortunately this factor does not have high enough predictive value to be used with confidence and, since it comes at the close of the training period of the prospective teacher, is of no apparent value in the selecting of those who should enter the professional training in education. Another general conclusion is that the attempt to devise tests to measure general teaching aptitude appears, in the light of the history of mental testing and present research results in this field, to hold greatest promise for the future. As yet the aptitude tests available are at best crude and of little value, although two or three are suggestive.

Hardesty (13) describes the following study. For a year following the issuance of a teaching credential to a teacher, the California State Department of Education secures a rating of success on the teacher. The teacher success ratings used as a basis for this study were secured from principals and supervisors on a rating scale on which command of subject matter, skill in instruction, and management and discipline were each rated numerically as excellent, good, fair, or poor. To determine the reliability of the rating, a second rating was secured on 127 of the subjects used in the study. The

---

13. Hardesty, Cecil F. Can Teaching Success be Rated? Nation's Schools 15:27-8, January 1935.

coefficient of correlation was .51 ± .03, 14 percent better than a chance guess. The grade point average and teaching success gave a correlation of .15. The correlation between grade point averages in professional subject matter and ratings on teaching success was .09. Grades in practice teaching and ratings in actual teaching gave a correlation of .07. The correlation of two ratings on command of subject matter was .05, on skill in instruction .02, and on management and discipline .05. When a study was made of grade point average correlated with ratings received on subject matter, a correlation of .25 ± .04 was found. This coefficient was low but significant and is 4 percent better than chance guess. Hardesty's conclusions are:

1. ratings given on success of teachers lack reliability.

2. grades received on courses in education and ratings received in practice teaching have so little relation to ratings given in classroom teaching that they are of no value in attempting to predict success of newly certified teachers.

At the University of Minnesota, a study has been going on for seven years regarding the Qualifying Examination (14). The examination measures the degree of preparation for assuming charge as student teachers. The purposes of the examination are:

1. to stimulate a more careful, thorough, and permanent preparation in the major teaching field and in the fundamental professional knowledge.

---

14. Qualifying Examinations in the College of Education at the University of Minnesota. School and Society, 35:698, May 21, 1932.

2. to furnish a basis for disqualifying or diverting by advisory means from student teaching, students of inadequate preparation.

3. to determine to what extent it is possible by such examination, in connection with other criteria, to predict successful performance as a teacher.

Corollary to the plan of examining students over the three principal fields of professional preparation an integrated nine-hour course covering the fields is to be instituted, beginning in the fall of 1932.

Bent and Douglass (15) in 1937 made a report on one phase of the qualifying examination, the relative number of failures on the professional examination of student teachers by departments at the University of Minnesota. The professional section of the examination which dealt with educational courses covered: a) educational psychology, b) secondary education, c) techniques of high school instruction or similar fields for elementary school teachers.

In Table I are the number of cases and the percentage of failures by major subject groups for the prospective teachers.

In their summary Bent and Douglass make the statement that it is quite possible that it does not require as great mental ability, command of English, and professional knowledge to "succeed" in physical education and vocational education as in teaching the academic subjects, but that that

---

15. Bent, R.K. and Douglass, H.R. Differences in the Performance of Departmental Groups of Student Teachers on Qualifying Examinations at the University of Minnesota. tabs. School and Society 45:726-7, May 22, 1937.

Table I

The Number of Cases and the Percentage of Failures by Major Subject Groups for the Prospective Teachers at the University of Minnesota

Major	Number of Students	Percent of Failures
Foreign Languages	58	0
Science	86	6
Social Studies	132	7
English and Speech	141	7
Educational and Liberal	147	10
Art and Music	134	12
Mathematics	26	15
Vocational Education	191	21
Nursing and Physical Education	171	29

position is not clearly incontrovertible and the differences found in this study raise some very interesting questions. In the present study it will be interesting to compare the ranking of these failures with the group rankings obtained on the Comprehensive Examination at Massachusetts State College.

Boardman and Patterson (16) in 1937 published further findings on this qualifying examination. Pre-examination prediction factors were:

16. Boardman, C.W. and Patterson, D. Achievement Examinations as a Technique for Selecting Students in Education. tabs School and Society, 49:586-8, May 6, 1939.

1. the Minnesota College Aptitude Test
2. the Miller Analogies Test
3. the Co-operative English Test
4. the Minnesota Reading Test
5. scholarship as measured by the honor point ratio

The examination consisted of four two-hour achievement examinations: high school work in the major subject, college work in the major subject, education, and English usage. The most important problem was to determine which of the various measures, singly or in combination, was of greatest significance in the selection of students. Two criteria of success are available: 1) scholarship as measured by the honor point ratio earned from all courses taken in the senior year, 2) success in student teaching as a measure of teaching ability as rated by several active supervisors. The method was to find the combination of pre-examination factors which had the highest relationship with the criterion and then to determine how much the examination would add to the coefficients. It was found that pre-examination factors involving previous scholarship were the best predictive factors. The additional information gained from the examinations is of no real value in predicting scholarship in the College of Education. The weight of evidence is against the examinations as a means of selecting students who will do better in student teaching. Among all the teaching fields here considered, it was found that the variable of previous scholarship, usually scholastic performance in the three years work in the major field, was a better predictive value than any of the achievement tests, whether those tests were "qualifying" examinations or were among the pre-examination factors.

Graduate Study: Concerning Graduate Study Lockwing (17) writes of the plan for individual study for the doctorate at Iowa. At the University of Iowa, when the doctor's "qualifying" examination is taken at least two full years before the conferment of the degree, and when the character of the qualifying examination demonstrates the candidate's competence and seriousness of purpose, permission to undertake individual study shall be granted.

Moore (18) gives graduate requirements for graduate study at Harvard. He states that it will probably long remain true that many students properly classed as graduate students will nevertheless require instruction identical with or similar to the more advanced undergraduates. The practice of counting graduate work for the Master's Degree by courses has now become distinctly out of date since the habit of giving the bachelor's degree on knowledge of a subject tested in a general examination has for most departments put an end to the successful scoring of courses as the sole method of proving fitness to graduate. Therefore the plan, proposed by the Department of Economics, and approved by the faculty, whereby the master's degree can be won only by examination on the subject, is a welcome change, and one which may well be considered by all departments.

---

17. Lockwing, Walter F. Individual Plan of Graduate Study at the University of Iowa. School and Society, 30: 743-5, November 30, 1929.

18. Moore, Clifford H. Requirements for Graduate Study at Harvard University. School and Society, 29:575-6, May 4, 1929.

Stumpf (19) in his study of graduate work in fifteen state teachers colleges mentions admission to candidacy and examination requirements. The candidate for a higher degree in all the institutions studied must demonstrate his fitness for advanced work. The announcements indicate that a knowledge of foreign language is not one of the important bases for admission to candidacy. (Note: degrees are given in various subject fields.) Four of the colleges make mastery of English a definite requirement. Qualifying examinations, when mentioned, are commonly optional; in other instances no reference is made to such examinations in the announcements. The announcements are not clear as to final requirements for the higher degree. Five colleges require either oral or written final examinations; one requires a demonstration of successful teaching; another requires a written final examination with an "oral" in the field of the thesis. The rest apparently are satisfied to accept course grades of "B" average or higher in full satisfaction of that portion of the degree requirements. Ten of the colleges require a thesis; in the others it is optional. The "major" professor or major department commonly supervises the study schedule and program of candidates for higher degrees, with frequent mention of the supervision given by a graduate committee or graduate council. The candidate ordinarily must present, for the approval of one or more of these agencies, a tentative program of work he intends to

---

19. Stumpf, W.A. Graduate Work in State Teachers' Colleges. tabs. School and Society, 46:834-8, December 25, 1937.

undertake, such presentation usually being required before admission to candidacy is granted. The question is raised as to whether, in a state teachers college, the advanced degree should be a "research" degree or a badge that designates superior teaching skill.

Comprehensive Examinations: Boyd (20) in writing of comprehensive examinations at the University of Kentucky, states that by requiring the comprehensives of all candidates for the degree - "we will give better direction and stimulus to the student and make possible a more unified and more vital grasp of the field and will affect favorably the methods of teaching for upper division students." The examination must test for fundamental concepts, for principles, for relationships, for ability to use knowledge in new problems, for real situations. Plenty of time must be given the student to think as well as to recall material committed to memory. An advantage claimed is that the student will be motivated to begin mastery of a field rather than to pass a succession of disconnected courses, that he will attempt education in its broadest, truest meaning. The elective system brought great good, but in destroying the comprehensive examination it rendered itself powerless to overcome its own weaknesses. No other proposal offers such large promise of integration, co-ordination, and motivation. The records at Harvard show that the level of achievement has been strikingly raised

---

20. Boyd, P.P. Compulsory Comprehensive Examinations. School and Society, 37:817-20, June 24, 1933.

(note: with tutoring) and the plan of the comprehensive examination for all is given the credit. The faculty's and administration's conduct of the plan is to be determined by trial.

Pilgram (21) claims that seniors at Franklin and Marshall will be led not merely to learn facts but also to think about these facts and to relate them over a wider period than is possible within the limits of a single course. In organizing their knowledge and in relating the various courses of the major, they will be guided by prepared outlines and reading lists, and by conferences with assigned members of the faculty.

In the annual report of Dr. Birkoff (22), acting dean of the Faculty of Arts and Sciences at Harvard, it is stated that few seniors lose their degrees because of failure to pass the general final examinations. The percentage of men failing to pass these examinations has shown a steady annual decline from about 10 percent to about 3.5 percent. Fewer were recommended for general honors. Perhaps it is to be regarded as a remarkable tribute to all concerned that about one undergraduate in three of those who took the general examinations should qualify for honors in special subjects.

Townsend (23) reports that the American Council on Edu-

---

21. Pilgram, Robert J. The Comprehensive Final Examination Plan Adopted for Franklin and Marshall College Seniors. School and Society, 33:60, January 10, 1931.

22. Birkoff, George D. General Final Examinations at Harvard College. School and Society, 43:86, January 18, 1936.

23. Townsend, M. Ernest. An Experiment in the Professional Examination of Teachers. School and Society, 50:537-41, October 21, 1939.

cation is attempting to aid a group of superintendents in improving their own selective procedures at the point of teacher employment. A battery of carefully constructed tests is to be prepared for administration to all candidates for teaching positions in the co-operating cities of the experiment. The co-operating cities will be asked to use as an important part of their teacher selection procedures adequate study of the "imponderables" and to weight these as of at least equal importance to the factors tested by the examinations. The tests are for the purpose of augmenting the other proper selective techniques. The tests are in three groups: 1) mental ability and general culture tests, 2) professional information tests, 3) special field tests. The Co-operative Test Service is responsible for the tests. One important factor in teaching success is that of intellectual competency in the professional field of specialization. In point of fact, documentary evidence already indicates that personal inadequacy, emotional insecurity, and lack of professional prestige frequently stem from an inadequate intellectual grasp of the nature of teaching and learning.

Corresponding to the experiment to be performed by the American Council on Education and reported by Townsend, a study was made at Massachusetts State College by Bracy (24) to ascertain general achievement of prospective teachers. He concludes that although prospective teachers in this college compare favorably in intelligence and in college marks with

---

24. Bracy, Alfred A. A Study of the General Achievement of Prospective Teachers at Massachusetts State College. Unpublished Master's Thesis, 1939.

other students in this college and in other colleges, they have not sufficient general education and grasp of special fields.

A very complete treatment of the comprehensive examination was made by Jones (25) in the form of an investigation for the Association of American Colleges.

From the above readings several important generalizations may be made:

1. Teacher Education is receiving more attention than ever before.

2. The problem of measuring or predicting teacher success seems paramount.

3. No single criterion appears adequate as a measure of teaching success.

4. The comprehensive examination has possibilities which should be investigated further.

The data from the study of the Comprehensive Examination at Massachusetts State College is found in Chapters IV and V.

---

25. Jones, Edward S. Comprehensive Examinations in American Colleges. New York: The Macmillan Company, 1933. pp xix + 436.

CHAPTER III

STATEMENT OF THE PURPOSE AND PROCEDURE

CHAPTER III

STATEMENT OF THE PURPOSE AND PROCEDURE

The Purpose Defined: The purpose of this study is to analyze the results of the Comprehensive Examination administered to Graduate Students in Education at Massachusetts State College with a view to its improvement and with the aim of discovering insofar as possible what it contributes toward a future measure of teaching success.

Subjects: The subjects are thirty graduate students who took the examination in May, 1939.

In the interests of analysis, an attempt was made to classify these subjects. These classifications used in the following analysis are:

1. graduates of various types of colleges
2. teachers active at the time of the examination and inactive, or prospective, teachers
3. those who received the master's degree in June, 1939 and those who did not for reason of lack of completion of the requirements for the degree
4. men and women

Materials: The materials are three test booklets for each of the thirty students. These booklets make up the examination.

The comprehensive examination consists of three parts which will be called hereafter, the "First Sitting", the "Second Sitting", and the "Third Sitting", respectively. To each of these sittings is allotted two hours of working time, making a total of six hours of working time for the whole examination.

The First Sitting was constructed by the Department personnel. It consists of four "sections", which are classified according to the four types of items used. Material from the following subjects appears in the First Sitting: History of Education, Principles and Methods of Teaching, Educational Psychology, Secondary Education, Classroom Management, Educational Tests and Measurements, School Law, Socialized Philosophy of Education, Educational Research and Statistics.

The material taken from these subject fields is placed in the examination in no sustained order. There are no boundaries between subject fields, the only divisions being according to the types of objective question used in the four sections.

The number of items of each type question are listed in Table II.

Table II

Number of Items Listed According to Type of Question in the First Sitting

Type	Number of Items
True-False	342
Multiple Choice	76
Completion	25
Matching	30
Total	473

The Second Sitting is a standardized educational test,

Newsom et al (26), for which tentative norms have been published.

The nature of the standardized test, as stated in the manual, is as follows:

"This examination is designed to measure students' abilities in the professional subject-matter of psychology and education, after they have completed the usually prescribed professional courses in preparation for teaching in secondary schools. It is not a measure of probable teaching success. On the other hand, it is designed to show a student's familiarity with and an understanding of the technical information and problems of education as they relate to the secondary school.

"The two forms (Form 1 and Form 2) of the test are the same length, essentially of the same difficulty and variability, and each contains the following broad fields of professional subject matter (sections): 1) Educational Psychology, 2) Tests and Measurements, 3) Principles of Organization and Management, and 4) Principles of Teaching.

"Even though the test items in each form are divided among the four fields mentioned, these fields have been considered from a very broad point of view. For example, under Educational Psychology will be found some material on child and adolescent psychology, mental hygiene, etc. Under Tests and Measurements, will be found material on the interpretation of elementary statistics and the philosophy of measurement. Principles of Organization and Management includes material dealing with social control, curriculum, guidance, school plant, professional relations, and school and community relations, records and reports, marks, etc. Included under Principles of Teaching are planning, types of teaching, and methods and techniques of teaching, and similar subjects."

The uses of the standardized test, as stated in the manual, are:

1. By teacher-training institutions to obtain a comprehensive survey of abilities resulting from professional training.

2. To select those students who have a sufficient fund of professional information and understanding of the problems of the secondary school to begin their practice teaching.

---

26. Newsom, N.W., Smeltzer, C.H., and Bowden, H.T. Comprehensive Examination in Secondary Education. Keokuk, Iowa: The Extra-Curricular Publishing Co., 1938. Forms 1 and 2.

3. As an instrument to measure progress in professional training; that is, it may be used as an initial and final test.

4. By local school officials for the selection of teachers.

5. To enable schools to compare the professional equipment of their students with the abilities of students in other schools."

The number of items on each section is listed in Table III.

Table III

The Number of Items on Each Section

of the Standardized Second Sitting

Section	Number of Items
Educational Psychology	88
Tests and Measurements	70
Principles of Organization and Management	75
Principles of Teaching	80
Total	313

The Third Sitting of the Comprehensive Examination was constructed by the Department personnel and is composed of 24 case studies in the fields of Supervision of Teachers, Secondary School Administration, Curriculum, Principles of Teaching, Guidance, and Classroom Management.

The emphasis is on the solving of actual or possible problems. Out of four to seven possible solutions to a problem, the student chooses as many as he thinks would give the

proper results in such a case. In scoring this sitting, deductions are made both for ignored correct solutions and for checked incorrect solutions. The purpose of this method of scoring is to correct both for lack of extensive knowledge and for guessing.

The Procedure: The procedure consists of two parts: 1) an analysis of the comprehensive examination itself, and 2) an analysis of the work done on the examination by the graduate students.

Part 1: Analysis of the Examination:

1. In any study which involves only 30 people, or 30 sets of observations, there is a question as to whether the measurements are reliable and significant. The first question to be considered is whether these tests are "reliable", i.e., if the tests were taken again by the same people, would they rank in the same order as on this first application? The reliability of the standardized Second Sitting was compared with the reliability given in the test manual, both being computed by the odd-even method.

There are two ways of obtaining the reliability of a test. One is to administer the same test twice, or to give two forms of the test to the same people, and correlate the scores. The resulting coefficient of correlation is the reliability. The other way is to correlate their scores on odd and even items respectively. This correlation, the reliability for half the test, may be converted to the reliability of the whole test by means of the Spearman-Brown formula.

Since in this examination, both Form 1 and Form 2 of the

standardized test were given, but to different people, it was necessary to compute the reliability for both forms. These reliabilities were checked with the manual.

2. The next step was to compute the reliability of the two locally-constructed First and Third Sittings.

3. The relationship which existed between each sitting and the Total Score, (the sum of the scores obtained on the three sittings), was obtained by correlating the score obtained on each sitting with the Total Score by the Pearson-Product-Moment method.

4. The relationship which each section of the standardized Second Sitting had with the Total Score was obtained by correlation.

5. The inter-relationships between the three sittings were obtained by correlation.

6. The total First Sitting score was correlated with each of the First Sitting sections to get their relationships.

7. The inter-relationships among the sections of the First Sitting were obtained.

8. An item analysis of the locally-constructed First Sitting was made to determine a statistical measure of the value of the items. This is not to be confused with "validity" of the items, i.e., the extent to which they measure what they are supposed to measure. The validity depends upon the experience of the local test makers. The approach here is toward statistical evaluation. An item was considered "good" which was answered correctly by less than 80 percent of the students and on which the average Total Score of those who

responded correctly to the item was greater than the average Total Score of those who responded incorrectly to the item. (An item is considered as "good" when the students with high scores answer it correctly.) The purpose of this item analysis is to revise the locally constructed sections of the examination. As the Third Sitting was composed of only 24 case studies, no analysis was made.

Part 2. Analysis of the Work Done on the Examination by Graduate Students:

1. Since the standardized Second Sitting has published tentative norms, the students who took the examination were compared in their knowledge of the basic courses in the field of Education with other students throughout the nation by means of "percentiles", a measure of relative standing. The percentile representing the average score on the separate sections was obtained.

The following basic questions will be asked regarding the students and the various classifications into which the students will be divided.

1. What is the relationship between scores on the Comprehensive Examination and success in college courses in Education?

2. What is the relationship between scores on the Comprehensive Examination and the number of years since graduation from undergraduate college?

3. What is the relationship between scores on the Comprehensive Examination and the type of undergraduate college attended?

4. What is the relationship between scores on the Comprehensive Examination and in-service versus prospective teachers?

5. What is the relationship between scores on the Comprehensive Examination and imminence of graduation in the year the examination was taken?

6. What is the relationship between scores on the Comprehensive Examination and sex of the students taking the Examination?

7. Are the high ranking students superior to the low ranking students in all parts of the Examination?

8. Are differences between groups significant?

Data obtained from the foregoing procedures will be found in Chapters IV and V under appropriate headings.

CHAPTER IV

COLLECTION AND INTERPRETATION OF DATA

Test Analysis

CHAPTER IV

COLLECTION AND INTERPRETATION OF DATA

Test Analysis

In this chapter data is collected and interpreted according to Part I of the procedure outlined in Chapter III, the analysis of the test itself. (Hereafter in any analysis in which the students are ignored as students and in which the Examination is considered only by itself the examination will be called the test.) This test will be analyzed by means of the responses to the items of the test without regard to the students or the various classifications into which they can be divided.

1. Test Results: The scores obtained on the whole test and on the three sittings appear in Table IV arranged according to rank of the Total Scores.

Table IV shows a great variation in scores obtained on the individual sittings and on the Total Score. The range of scores compared to the maximum score obtained on each sitting and the Total are: First Sitting: 206 points range in 283; Second Sitting: 140 points in 216; Third Sitting: 49 points range where the maximum positive score was 35 (and the maximum negative score was -14); Total Score 387 points in 534.

2. Conversion of Scores: Since two forms of the standardized sitting were given to different people (11 of whom took Form 1, and 19 Form 2) and since the two forms are not exactly equal in difficulty, it was necessary to convert the

Table IV

Scores Obtained on the Whole Examination and the Three Sittings

Student	First Sitting	Second Sitting	Third Sitting	Total Score
JAF	283	216	35	534
BRL	229	197	32	458
MTC	221	188	23	432
HCH	216	184	20	420
GEM	239	161	15	415
SGG	217	193	2	412
UCC	212	170	5	387
H G	189	164	4	357
APG	181	157	18	356
JMB	180	157	13	350
MMC	181	155	10	346
CRG	202	140	4	346
HRS	208	138	-11	335
FTD	176	128	28	332
MAC	173	135	17	325
RAS	180	135	0	315
SMG	163	142	2	307
ALS	171	140	-4	307
WPM	148	125	28	301
RHG	134	154	8	296
RAP	163	131	1	295
RPH	172	122	-14	280
EJW	137	110	20	267
HKR	122	139	-5	256
ADC	150	93	6	249
WJC	132	103	-3	232
FJC	120	112	-10	222
JEG	115	97	-4	208
RWV	90	123	-10	203
LDB	77	76	-6	147

scores to the same basis by means of percentiles. (A "percentile" of 90 means that the individual obtained a score equal to or better than 90% of the people who took the test.) Thus, a score of 164 on Form 2 falls on the 60th percentile. The score represented by the 60th percentile on Form 1 is 174. Therefore a score of 164 on Form 2 corresponds to a score of

174 on Form 1. All the scores on Form 2 were converted to their corresponding scores on Form 1 in this manner.

The scores obtained on each section of the standardized Second Sitting appear in Table V. The scores obtained on the standardized Second Sitting also show a great variation. The range of scores compared to the maximum score obtained on each section and the total are: Educational Psychology: 48 points range in 67; Tests and Measurements: 35 points in 49; Principles of Organization and Management: 23 points in 49; Principles of Teaching: 28 points in 54; total score for the sitting: 140 points in 216.

3. Reliability of the Standardized Second Sitting: As 11 people took Form 1 of the standardized Second Sitting and 19 took Form 2 which used none of the items present in Form 1, it was necessary to obtain the reliability of each form separately. This was done by the odd-even method, as explained in Chapter II. Correlation was computed by the Spearman Rank Differences Method.

The reliability for half Form 1 was .82; for half Form 2: .85. The reliabilities corrected respectively by the Spearman-Brown formula (1) for reliability of each total form are given in Table VI and compared with reliabilities of the two forms given in the test manual. The reliability as given in the manual was computed on the basis of 100 people for each form. In this study the Probable Error is used as a measure of variability with coefficients of correlation.

---

1. Appendix: The Spearman-Brown Formula

Table V

Scores Obtained on the Second Sitting and the Four Sections of the Sitting

Student	Part 1 Educational Psychology	Part 2 Tests and Measure-ments	Part 3 Principles of Org. and Man.	Part 4 P Principles of Teaching	Total
JAF	67	49	49	52	216
BRL	67	37	44	48	197
MTC	54	40	40	54	188
HCH	50	38	42	52	184
GEM	50	21	49	41	161
SGG	44	53	41	54	193
UCC	62	29	37	42	170
H G	49	31	47	37	164
APG	41	43	29	44	157
JMB	39	21	41	52	157
MMC	41	34	40	40	155
CRG	39	25	44	33	140
HRS	34	25	48	33	138
FTD	34	14	41	39	128
MAC	41	30	32	35	135
RAS	44	16	41	39	135
SMG	36	24	36	46	142
ALS	34	31	28	47	140
WPM	28	19	36	40	125
RHG	52	28	35	39	154
RAP	28	35	35	33	131
RPH	38	34	22	28	122
EJW	20	19	40	29	110
HKR	49	15	39	36	139
ADC	21	14	29	30	93
WJC	20	23	33	26	103
FJC	26	20	36	29	112
JEA	25	18	21	33	97
RWV	28	29	38	24	123
LDB	19	10	26	26	76

From Table VI it is evident that the reliabilities as calculated on the basis of 11 and 19 people respectively check with a sifficient degree of accuracy the reliabilities of the two forms as given in the test manual. It can be said that the people who took the test would rank in almost the same order on another application of the test.

Table VI

The Reliability of Form 1 and Form 2 of the Second Sitting as Computed, and as Given in the Test Manual

Number of Students at Mass. State	Reliability Form 1	Number of Students in Manual	Reliability Form 1
	Form 1		
11	.90 ± .05	100	.90
	Form 2		
19	.92 ± .03	100	.89

The Index of Reliability (2), which shows the consistency with which the test measures what we are measuring, was .95 for Form 1 and .96 for Form 2.

4. Reliability of the First and Third Sittings: The reliability for the First Sitting on the basis of 30 people was .95 ± .01. The reliability of the Third Sitting, also on the basis of 30 people, was .78 ± .05.

The Indices of Reliability were .96 and .88 respectively.

From these reliability figures it can be seen that the First and Second Sittings are very reliable. The Third Sitting is not very reliable in its present length. Because of the fact that in the Second Sitting both forms were administered using different items in the two forms, it was considered impossible to find the reliability of the whole test.

5. Correlation of Total Score with Each Sitting: The relationship which existed between each sitting and the Total

---

2. Appendix: Index of Reliability

Table VII

Relationships between Total Score
and the Score on Each Sitting

Total Score with First Sitting:	.96 ± .01
Total Score with Second Sitting:	.94 ± .02
Total Score with Third Sitting:	.67 ± .07

Score (the sum and the scores obtained on the three sittings) was obtained by correlating by the Pearson Product-Moment Method the scores obtained on each sitting with the Total Score.

Sections of an examination are considered "good" when they correlate highly with the criterion and lowly with each other. By correlating highly with the criterion they show that they are a good measure of the criterion. By correlating lowly with each other they show that they are not measuring the same knowledge or, perhaps, the same ability.

The First and Second Sittings correlate highly with the Total Score which is used as the criterion. The Third Sitting does not correlate highly with the criterion. It is therefore open to question as to its value in the Examination. Its outstanding fault is the relatively few number of items contained in it. The range of scores obtained on the Third Sitting (1. Test Results) was about one-third that of the Second Sitting and about one-fourth that of the First Sitting. Further data may show the value of this sitting. If so, a recommendation will be made as to its length and scoring.

6. Correlation of Total Score and Each Section of the Second Sitting: The relationships which existed between each section of the standardized Second Sitting were obtained by correlation and appear in Table VIII.

Table VIII

Relationships between Total Score and Each Section of the Standardized Second Sitting

Total Score with Educational Psychology:	.82 ± .04
Total Score with Tests and Measurements:	.68 ± .07
Total Score with Principles of Organization and Management:	.65 ± .07
Total Score with Principles of Teaching:	.81 ± .04

All these relationships are highly significant (3). Of these sections, Educational Psychology and Principles of Teaching have the highest relationships with the Total Score, that is, with the whole test.

7. Inter-relationships among the Sittings: The inter-relationships among the three sittings were obtained by correlation, and are listed in Table IX.

Table IX

Inter-relationships Among the Three Sittings

First Sitting with Second Sitting:	.84 ± .04
First Sitting with Third Sitting:	.54 ± .09
Second Sitting with Third Sitting:	.54 ± .09

3. Appendix: Significance of Correlations

Having found that both the First and Second Sittings had a high relationship to the criterion (5. Correlation of Total Score with Each Sitting) and that therefore they are a good measure of the criterion, there now remains to be seen the degree to which they tend to measure the same knowledge or abilities. That information will be found in the inter-correlations (Table IX). It was found that the First and Second Sittings correlate, .84 ± .04, which is a relatively high relationship. The First Sitting and the Second Sitting measure to too high a degree the same knowledge or abilities. Under "Material" it is noticeable that they cover to some extent the same courses. The Third Sitting again involves the lowest relationships as mentioned in (5).

8. Correlations between First Sitting Score and Section Scores: In order to obtain the degree of relationship between each of the sections of the First Sitting and the score on that sitting, the total scores on that sitting were correlated with the individual section scores. The results appear in Table X.

Table X

Relationships between First Sitting Score and First Sitting Section Scores

First Sitting Score with True-False:	.97 ± .01
First Sitting Score with Multiple Choice:	.76 ± .05
First Sitting Score with Completion:	.56 ± .09
First Sitting Score with Matching:	.64 ± .08

Using the same method of attack as in 5. and 7. and using the First Sitting score as the next criterion it is found that the True-False and Multiple Choice sections correlate most highly with the First Sitting score. The Completion and Matching sections could well be omitted or increased in length, as they do not correlate highly with the criterion.

9. First Sitting Inter-correlations: The inter-relations among the sections of the First Sitting are shown in Table XI.

Table XI

Inter-relationships among the Sections of the First Sitting

True-False with Multiple Choice:	.64 ± .07
True-False with Completion:	.43 ± .10
True-False with Matching:	.51 ± .09
Multiple Choice with Completion:	.53 ± .09
Multiple Choice with Matching:	.50 ± .09
Matching with Completion:	.56 ± .09

There now remains to be seen whether the True-False and Multiple Choice sections measure the same thing. The inter-correlation between them (Table XI) is .64 and is significant. This correlation is no higher than the better correlations sometimes found between intelligence and various course marks, so it is decided that the True-False and Multiple Choice sections, from this viewpoint, are worth retaining.

All the correlations listed in Table XI are highly sig-

nificant except the correlation between True-False and Completion questions. That correlation however is significant.

10. Item Analysis: For the purpose of revision of the locally constructed First Section, an item analysis was made according to the method listed under "Procedure", Chapter III. A "good" item was considered to be an item which was answered correctly by from 20% to 80% of the students taking the test, and on which the average score (on the Sitting) of those answering the item correctly was higher than the average score of those answering the item incorrectly. The results appear in Table XII.

Table XII

Item Analysis of the First Sitting

Total number of items on the First Sitting:	473
Items rejected for reason of being too easy:	98
Items rejected for reason of being too difficult:	55
Items rejected for reason that students who obtained low scores answered them better than did students who obtained high scores:	55
Total number of items rejected:	208
Total number of items retained:	265

Examples of rejected items:

1. True-False item #19 was answered correctly by 28 students out of 30. It was rejected for the reason that it was too easy.

2. True-False item #75 was answered correctly by only 1 student out of 30. It was rejected for the reason that it was too difficult to hold a place in this test.

3. The average score obtained on this sitting by those who answered True-False item #16 correctly was 164.75. The average score of those who answered the item incorrectly was 181.79. The poorer students answered this item better than the better students, which would lead us to question either the item or its wording. The item was rejected.

In the revision of the test this item analysis will be tempered with the judgment of those constructing the test with regard to the validity of rejected or questionable items. The above item analysis should serve as an aid to judgment, not as an absolute criterion.

The collection and interpretation of data which has a greater concern for the people who took the examination appears in Chapter V.

CHAPTER V

COLLECTION AND INTERPRETATION OF DATA:

Examination Analysis

CHAPTER V

COLLECTION AND INTERPRETATION OF DATA:

Examination Analysis

In this chapter data is collected and interpreted according to Part 2 of the procedure outlined in Chapter III: analysis of the work done on the Examination by graduate students.

1. Percentiles Attained on the Second Sitting: Since there are published tentative norms for the standardized Second Sitting, the students who took the Examination were compared in their knowledge of the basic courses in the field of Education with other students throughout the nation. The percentiles which correspond to the Second Sitting scores of Table V are listed in Table XIII.

The range of percentiles obtained on each section of the sitting are as follows: Educational Psychology, 1 to 88; Tests and Measurements, 1 to 98; Principles of Organization and Management, 1 to 88; Principles of Teaching, 3 to 88; total, 1 to 94. This standardized sitting was designed for students who had taken four or more basic courses in the fields of Education and Psychology. It seems reasonable to assume that in order to take several courses in a subject field the student must have been able to pass successfully the requirements for each of the courses. If then, it is considered that the percentiles given in the manual are representative of the quality of work done by people who were able to pass successfully the requirements of these courses, it may be said that

Table XIII

Percentiles Attained on the Standardized Second Sitting and the Four Sections of the Sitting

Student	Part 1 Educational Psychology	Part 2 Tests and Measurements	Part 3 Principles of Org. and Man.	Part 4 Principles of Teaching	Total
JAF	88	95	88	80	94
BRL	88	70	72	67	82
MTC	46	80	53	87	74
HCH	33	75	64	80	70
GEM	33	16	88	40	41
SGG	20	98	60	88	78
UCC	74	44	40	44	55
H G	31	50	81	27	48
APG	13	18	59	80	37
JMB	15	87	10	52	37
MMC	15	60	53	38	35
CRG	13	30	72	16	22
HRS	7	30	86	16	20
FTD	7	4	58	34	14
MAC	15	47	20	19	18
RAS	20	7	59	33	18
SMG	9	27	35	59	23
ALS	7	50	8	62	22
WPM	5	11	33	38	11
RHG	40	39	29	35	33
RAP	4	63	29	16	15
RPH	12	60	2	7	9
EJW	1	11	55	8	4
HKR	30	5	49	23	21
ADC	1	4	10	10	1
WJC	1	23	21	5	3
FJC	3	13	33	9	5
JEG	3	10	1	15	2
RWV	4	43	42	3	10
LDB	1	1	5	5	1

these graduate students are distributed, at the time of taking the Examination, from a level of mere passing grade to a level which exceeds 94% of the students in undergraduate education courses throughout the nation.

2. Second Sitting Averages: The average score obtained on each of the sections and the total score of the Second

Sitting with the corresponding percentiles are listed in Table XIV.

Table XIV

The Average Score and Corresponding Percentiles for Each Section and Total Score of the Standardized Second Sitting

Section	Subject Field	Average Score	Percentile
1	Educational Psychology	39.33	13
2	Tests and Measurements	27.50	39
	Principles of		
3	Organization and Management	37.30	41
4	Principles of Teaching	38.70	33
	Total	142.83	24

Table XIV is to be read: the student who represents the average of our graduate students exceeds in Educational Psychology 13% of the people throughout the nation who have taken this sitting; in Tests and Measurements 39%; in Principles of Teaching 33%; in the Total Score: 24%.

3. Conclusions: The average student who took the Examination did not score as highly on these sections as the average student throughout the nation. The average student is particularly weak in Educational Psychology.

4. Questions: In this chapter the study is concerned with certain basic questions regarding the groups who took the Examination. These questions will follow.

5. The first basic question to be answered is: "What is the relationship between scores on the Comprehensive Examination and success in college courses in Education?"

Regarding the Examination, the question naturally arose as to whether the examination measured the same thing or something different than the grades obtained in professional courses. The data in answer to this question is found in succeeding sections.

A. Courses and Grades: The number of courses studied by each student in the field of Education in the Graduate School at Massachusetts State College, and the average grade received in these courses are listed in Table XV. Average grades received in Education courses vary from 72 to 92.

B. Correlation of Average Grades and All Scores: The relationships between Average Grades and the various scores on the Examination are given in Table XVI.

The highest relationship (between Average Grades and scores) exists between Average Grades and the standardized Second Sitting, which shows that the Second Sitting resembles course' examinations and work more than do the locally constructed sittings. The three lowest relationships were between Average Grades and the Third Sitting, Tests and Measurements, and Principles of Organization and Management respectively, in decreasing order. Out of 30 students, only 15 have

Table XV

The Number of Courses Studied in the Graduate School at Massachusetts State College in the Field of Education, and the Average Grade Received in Those Courses

Student	Number of Courses	Average Grade	Student	Number of Courses	Average Grade
JAF	8	86.37	RAS	3	84.33
BRL	7	86.43	SMG	3	81.00
MTC	6	90.33	ALS	3	85.33
HCH	5	91.60	WPM	6	82.00
GEM	3	85.33	RHG	4	90.00
SGG	5	87.00	RAP	3	79.00
UCC	8	88.25	RPH	5	82.60
H G	4	87.50	EJW	7	82.43
APG	3	82.66	HKR	3	82.33
JMB	6	81.66	ADC	3	79.66
MMC	3	82.33	WJC	4	85.75
CRG	4	79.25	FJC	5	75.40
HRS	6	80.83	JEG	3	76.00
FTD	7	82.00	RWV	4	71.75
MAC	1	77.00	LDB	9	76.22

Table XVI

Correlation Between Grades in Education Courses Studied in the Graduate School at Massachusetts State College and Scores on the Examination

Average Grades with Total Score:	.67 ± .07
Average Grades with First Sitting:	.62 ± .08
Average Grades with Second Sitting:	.69 ± .07
Average Grades with Third Sitting:	.43 ± .10
Average Grades with Educational Psychology:	.66 ± .07
Average Grades with Tests and Measurements:	.42 ± .10
Average Grades with Principles of Organization and Management:	.34 ± .10
Average Grades with Principles of Teaching:	.67 ± .07

had a course in Tests and Measurements at any time. The Organization and Management coefficient is not significant. The conclusion with regard to the Third Sitting is that of the three sittings that sitting corresponds least to education examinations and work. It was designed as a thought-provoking section and appears to be somewhat better in that respect than other sections or sittings.

C. Correlation of the Number of Education Courses Taken and the Total Score: The relationship between the number of graduate education courses taken and the Total Score on the Examination is .21 ± .12 which is not significant, and shows that with an increase in the number of graduate education courses taken, there is practically no probability of an increased score on the Examination. Because of the low magnitude of this coefficient of correlation, it was felt that no further correlations need be made with the number of education courses taken as one factor of the correlations.

D. Conclusions: The answer to the question of relationship between Examination scores and success in college courses in Education is found to be that there is no higher relationship than is usually found in the best correlations obtained between intelligence and marks in courses, or between marks in different subjects.

The answers to the question as to whether the Examination measured the same thing or something different than the grades obtained in professional courses are:

1. While the Examination necessarily measures some of the same things as course grades, the correlation of .67 for the whole Examination would indicate that the Examination does not measure to a high degree the

same thing as course marks. Therefore it would appear to be worthwhile to retain the Examination as a supplementary device for measuring the students' grasp of the general field.

2. The standardized sitting measures to a higher degree than the locally constructed sittings the same thing as course marks.

3. The Third Sitting, composed of case problems, measures to a much smaller degree the same thing as course marks. It would appear that in the solving of case problems, where an effort has been make to force the student to think in terms of educational principles, the results have a low relationship to college marks.

6. The second basic question to be answered is: "What is the relationship between scores in the Examination and the number of years since graduation from undergraduate college?"

A. Correlation of Years Out of College and Scores on the Examination: Figures representing the number of years out of college for each graduate student were correlated with all scores on the Examination. The results appear in Table XVII.

B. Conclusions: None of the correlations in Table XVII is significant. The conclusion is drawn that for these students there is no significant relationship between the number of years out of college and any sitting or section of the Examination. The controlling element in these correlations was the large group of students who had been out of college less than six years, and who because of their great variability, overshadowed in correlations the few students who had been out of college from 6 to 28 years. It would be interesting to obtain the same relationships between number of years out of college and scores on the Examination for students who were equally divided in the same 14 two year groups.

Table XVII

Correlations of the Number of Years Out of College for Graduate Students with Scores on the Examination

Number of years out of college with Total Score:	-.04 ± .12
Number of years out of college with First Sitting:	.06 ± .12
Number of years out of college with Second Sitting:	-.04 ± .13
Number of years out of college with Third Sitting:	-.11 ± .13
Number of years out of college with Educational Psychology:	.36 ± .11
Number of years out of college with Educational Tests and Measurements:	-.14 ± .12
Number of years out of college with Principles of Organization and Management:	.12 ± .12
Number of years out of college with Principles of Teaching;	-.12 ± .12

7. The third basic question is: "What is the relationship between scores on the Comprehensive Examination and the type of undergraduate college attended?"

A. Average Score and Average Percentiles on the Standardized Second Sitting for Graduates of Different Type Colleges: The average score on the Second Sitting and the corresponding percentile was obtained for each of the five types of undergraduate colleges from which these students graduated. These averages appear in Table XVIII.

Percentiles range from 4 to 94. The percentile corresponding to the average score on the standardized sitting was previously listed (Table XIV) as 24. In Table XVIII the Engineering student, who did all his work in education at this college, attained a percentile of 94. From this percentile, the conclusion is drawn that it is possible to do a quality of

Table XVIII

Average Score on the Second Sitting and Corresponding Percentile for Graduates of Different Type Colleges

Type College	Number of Students	Average Second Sitting Score	Percentile
Engineering	1	216.00	94
Liberal Arts	19	148.95	28
Art	2	147.00	27
Physical Education	3	132.00	16
Teachers' College	5	109.80	4

work in this department that ranks high on national norms. The reason, then, for a comparatively low general average percentile and low group average percentiles must be attributed either to the students' capability or to other factors. As many factors as possible will be investigated in this study.

It is interesting to compare the rank of the groups in Table XVIII with the rank of similar groups arranged from the study made by Bent and Douglass (15) Table 1, on Page 7, Chapter I. In ths group of Teachers' College students, the Educational and Library and the Vocational Education groups are excluded for the reason that they are not comparable with groups in the present study. When the remaining groups are combined by subject majors to give new groups of Liberal Arts, Art, and Physical Education, the results appear in Table XIX.

The rank in the present study is in exactly the same order as for these Teachers' College students classified in corresponding groups.

B. The Average Total Score for Graduates of Different Type Colleges: The average Total Score was obtained

Table XIX

Table I Re-arranged into New Classifications

Major	Number of Students	Percent of Failures
Liberal Arts	441	7
Art	134	12
Physical Education	171	29

for each of the five types of colleges from which these students graduated. These averages are listed in Table XX.

Table XX

Average Total Score for Graduates of Different Type Colleges

Type College	Number of Students	Average Total Score
Engineering	1	534.00
Liberal Arts	19	336.79
Art	2	301.00
Physical Education	3	284.33
Teachers College	5	260.40

The types of colleges have the same rank in both Table XVIII and Table XX, the standardized sitting and the Total Score, and they both agree, in corresponding classifications, with the study be Bent and Douglass mentioned in step 9, this chapter.

C. Average Scores on the First Sitting for Graduates of Different Type Colleges: The average scores on the First Sitting obtained by each of the five types of colleges appear in Table XXI.

Table XXI

Average First Sitting Scores for Graduates of Different Type Colleges

Type College	Number of Students	Average First Sitting Scores
Engineering	1	283.00
Liberal Arts	19	179.68
Physical Education	3	159.33
Art School	2	151.50
Teachers College	5	140.60

On the First Sitting the Physical Education and Art groups interchanged places in rank order.

D. Average Scores on the Third Sitting for Graduates of Different Type Colleges: The average scores on the Third Sitting obtained by each of the five groups of graduates of different type colleges appear in Table XXII.

The first real change in rank of these five types of college graduates appears on the sitting which is composed of case study problems. The Teachers' College group moved up to second place displacing the Liberal Arts, Art, and Physical Education groups which still retained their relative rank. With so few people in each group the reliability of differences

Table XXII

Average Third Sitting Scores for Graduates of Different Type Colleges

Type College	Number of Students	Average Third Sitting Scores
Engineering	1	35.00
Teachers' College	5	10.00
Liberal Arts	19	8.16
Art School	2	2.50
Physical Education	3	-7.00

is not considered here, but it would appear that the Teachers' College group might have some relative advantage on these case situations.

E. Average Grades for Graduates of Different Type Colleges: The average grades received in education courses studied in the Graduate School of Massachusetts State College appear in Table XXIII in order of rank.

Table XXIII

Average Grades for Graduates of Different Type Colleges in Education Courses Studied in the Graduate School at Massachusetts State College

Type College	Number of Students	Average Grade
Engineering	1	86.37
Liberal Arts	19	83.85
Art	2	82.33
Teachers' College	5	80.26
Physical Education	3	78.97

The rank order of grades agrees essentially with the rank order of scores received on the standardized Second Sitting and the Total Score with the exception that the lowest two groups are interchanged in rank.

F. Conclusions: In the present study there are too few cases on which to predict what graduates of different type colleges will do on future examinations. However, if this classification is regarded as a grouping of individuals, rather consistent results are obtained. On the Total Score, the three sitting scores, and average grades, the only real upset in rank of these five groups occurs on the Third Sitting case problems on which the Teachers' College group did its best comparative work. The results on this question would lead to the conclusion that student effectiveness on the first two sittings was due to general ability, while on the Third Sitting effectiveness for the Teachers' College group was due either to special training or special ability, while general ability still held control for the other four groups of students.

8. The fourth basic question to be answered is: "What is the relationship between scores in the Comprehensive Examination and in-service versus prospective teachers?"

A. Average Scores and Average Percentiles on the Standardized Second Sitting for Active and Inactive Teachers: On the Second Sitting the average score for 17 active teachers was 143.47. The average score for 13 inactive teachers was 142.00. The percentiles are 25 and 24 respectively. The reliability of differences will be found in step number 12.

The above difference is not reliable. Of the 13 inactive teachers, 12 were full time graduate students at this college.

B. Scores for Active and Inactive Teachers: The average Total Score, the average sitting scores, and the average scores on the sections of the standardized sitting for active and inactive teachers are listed in Table XXIV.

The only section on which the inactive teachers exceeded the active teachers was that of Tests and Measurements. It is of interest that of the inactive teachers, 10 out of 13 had studied a course in Tests and Measurements at some time in their college careers, while of the active teachers only 5 out of 17 had studied such a course.

The differences in Table XXIV are not reliable but it is noticeable that they are consistent.

C. Percentiles on the Sections of Standardized Second Sitting for Active and Inactive Teachers: The percentiles which correspond to the average scores of Table XXIV are listed in Table XXV.

Percentiles range from 10 to 44. On Tests and Measurements, the percentiles have a greater difference than the scores obtained on the same section due to a steeply sloped percentile curve.

D. Average Grades for Active and Inactive Teachers: The average grade received in education courses studied in the Graduate School at Massachusetts State College by active teachers was 84.27; by inactive teachers, 80.75. This difference is in the proper direction to coincide with all their differences on the Examination except the section on Tests and Measurements, in which so few active teachers had taken

Table XXIV

Scores for 17 Active and 13 Inactive Teachers

Part of the Examination	17 Active Teachers	13 Inactive Teachers
Total Score	325.65	319.54
First Sitting	173.12	172.15
Second Sitting	143.47	142.00
Third Sitting	9.06	5.38
Sections of the Second Sitting		
Educational Psychology	41.12	37.00
Tests and Measurements	26.29	29.07
Principles of Organization and Management	37.82	36.61
Principles of Teaching	38.77	38.61

Table XXV

Percentiles on the Sections of the Standardized Second Sitting Attained by Active and Inactive Teachers

Section	Active Teachers	Inactive Teachers
Total Score on the Sitting	24	24
Educational Psychology	15	10
Tests and Measurements	35	44
Principles of Organization and Management	43	38
Principles of Teaching	33	33

such a course. The difference is not reliable.

E. Conclusions: In general, the active teachers surpassed the inactive or prospective teachers and obtained

higher grades in courses. Due to the limited number of cases, differences are not reliable. In the one section in which the active teachers were surpassed, a majority had not studied the subject matter used as the basis for the section.

9. The fifth basic question to be answered is: "What is the relationship between scores on the Examination and imminence of graduation in the year the Examination was taken?"

A. Average Scores and Average Percentiles on the Standardized Second Sitting for Those Who Received the Master's Degree and for Those Who Did Not: The average score made on the standardized sitting by 15 students who received their Master's degrees in June 1939 was 140.40. The average score made by those 15 students who did not receive their degrees was 145.27. The corresponding percentiles were 22 and 26 respectively. The difference was not reliable.

B. Scores for Students Who Received Their Degrees and for Those Who Did Not: All of the average scores for those students who did and did not receive their degrees are listed in Table XXVI.

The students who did not receive their Master's degree at the end of the year exceeded the Masters in all sections of the Examination except in Principles of Organization and Management, in which the Masters exceeded the other group by a slight margin. The differences were not reliable.

C. Percentiles on the Second Sitting for Masters and Bachelors: The percentiles which correspond with the average scores of Table XXVI are listed in Table XXVII.

Table XXVI

Scores for 15 Students Who Received Master's Degrees and 15 Students Who Did Not Receive the Degree

Part of the Examination	15 Masters	15 Bachelors
Total Score	306.60	339.40
First Sitting	164.60	187.47
Second Sitting	140.40	145.27
Third Sitting	1.60	13.33
Sections of the Second Sitting		
Educational Psychology	39.27	39.40
Tests and Measurements	27.27	27.73
Principles of Organization and Management	38.20	36.40
Principles of Teaching	35.87	41.53

Table XXVII

Percentiles on the Sections of the Standardized Second Sitting Attained by Students Who Received Their Degrees and by Students Who Did Not Receive the Master's Degree

Section	Masters	Bachelors
Total Score on the Sitting	22	26
Educational Psychology	13	13
Tests and Measurements	38	39
Principles of Organization and Management	45	37
Principles of Teaching	22	43

The Masters exceeded the Bachelors only in Principles of Organization and Management. Differences in scores are not reliable. The differences are magnified on the percentile curve.

D. Average Grades for Students Who Did and Did Not Receive the Master's Degree: The average grade of the students who did receive the degree was slightly higher than that of those students who did not receive the degree: 83.48 as compared to 82.01.

Between these two groups the differences in grades and the differences in Examination scores are not in the same direction.

E. Conclusions: The Masters and Bachelors establish a doubtful point. The Masters obtained a slightly higher average of grades, and a lower average score on the Total Score, each sitting, and each section except Organization and Management. In order to explain this first discrepancy in the consistency of results the following lines of thought are advanced:

1. Differences are not reliable.

2. Those who received the degree may have a better concept of Organization and Management which quality may have aided in their acquiring the degree. This might explain their one relative success in Organization and Management.

3. The quality of graduate work or of graduate students may be improving.

10. The sixth basic question is: "What is the relationship between scores on the Comprehensive Examination and sex of the students taking the Examination?"

A. Average Scores and Average Percentiles on the Standardized Second Sitting for Men and Women: On the Second Sitting the average score of 20 men was 142.20. The average score of 10 women was 144.10. The corresponding percentiles are 24 and 25 respectively. Differences are not significant.

B. Scores for Men and Women: The average scores on the parts of the Examination appear in Table XXVIII.

Table XXVIII

Scores for 20 Men and 10 Women

Part of the Examination	20 Men	10 Women
Total Score	320.10	328.80
First Sitting	170.25	176.40
Second Sitting	142.20	144.10
Third Sitting	7.05	8.30
Sections of the Second Sitting		
Educational Psychology	39.35	39.30
Tests and Measurements	26.90	28.70
Principles of Organization and Management	36.00	36.90
Principles of Teaching	37.35	41.40

The women led the men on all parts of the Examination with the exception of the Educational Psychology section on which the scores were practically equal. The differences in Table XXVIII are not reliable, but once again it may be said that differences are consistently in favor of one group; in this case, the women.

C. Percentiles on the Standardized Second Sitting for Men and Women: The percentiles which correspond to the average scores in Table XXVIII are listed in Table XXIX.

Once again differences between the groups are exaggerated by percentiles. One point difference in score is equivalent to three or four points difference in percentiles on the steep slope of the curve. The range of percentiles listed in the table for men and women is 13 to 43.

Table XXIX

Percentiles on the Sections of the Standardized Second Sitting Attained by Men and Women

Section	20 Men	10 Women
Total Score on the Sitting	24	25
Educational Psychology	13	13
Tests and Measurements	37	43
Principles of Organization and Management	35	39
Principles of Teaching	28	42

D. Average Grades for Men and Women: The average grade received by men in education courses studied in the Graduate School was 82.87. The average grade for women was 82.50. This slight difference is not reliable and is in the opposite direction from the differences in scores made on the Examination.

E. Conclusions: The study made of the grades and scores made by men and women furnishes the second example of

a slight difference in grades being in the opposite direction to the difference in Examination scores. While the slight difference in grades is not reliable, it makes a slight exception to the rather remarkable consistency of results. The men who took the Examination were more variable than the women in that they scored higher and lower than the women.

11. The seventh basic question is: "Are the high ranking students superior on all parts of the Examination?"

A. Comparison of the Average Scores Made by the 5 Ranking Students with Average Scores Made by the Lowest 5 Students: The average scores made on all sittings and sections of the Examination by the 5 ranking students are compared with average scores made by the lowest 5 students in Table XXX.

Table XXX

Average Scores Made by the 5 Ranking Students and by the 5 Lowest Students

Part of the Examination	Average Score for 5 Highest Students	Average Score for 5 Lowest Students
Total Score	451.80	202.40
First Sitting	237.60	106.80
Second Sitting	189.20	102.20
Third Sitting	25.00	-6.60
Sections of the Second Sitting		
Educational Psychology	57.60	23.60
Tests and Measurements	37.00	20.00
Principles of Organization and Management	44.80	30.80
Principles of Teaching	49.40	27.60

Table XXX includes the work of 10 people of the total 30. As there is contained in each group the work done by only 5 people, the significance of the differences was not calculated. It is noticeable that the upper and lower groups are consistent in their relative scores.

B. Average Grades for the Highest 5 and Lowest 5 Students: The 5 students who ranked highest on the Examination obtained an average grade, for courses taken in the Graduate School, of 88.01; the average grade for the lowest 5 students was 77.02. Again, due to the few people studied, the reliability of the difference was not calculated.

C. Conclusion: It is apparent that the highest 5 students exceeded the lowest 5 in each section of the examination and in average grades received in the Graduate School.

12. The eighth basic question is: "Are differences between groups significant?"

A. Significance of Differences: In any study which involves averages of scores for different groups of people, the question arises as to the "significance" or "reliability" of these differences, i.e., if the study were repeated with the same materials and with similar subjects, would the same groups again exceed the groups which were found to be relatively poorer in the first study, or, in the second study, would the situation be reversed such that the groups formerly considered poorer would now exceed? This question is answered in most studies by the calculation of the "Critical Ratio".

B. Critical Ratios for the Data: As there are so few people in this study, it was thought advisable to start

calculating critical ratios where the number of students involved would be as large as possible, namely in the classifications in which there were two groups nearly equally divided. It was found that with this limited number of people, differences were not significant. For example: 17 active and 13 inactive teachers scored on Tests and Measurements 26.29 and 29.07 points respectively. The critical ratio was .69. Since a value of 3.00 is the least value of the critical ratio which can be used as a criterion for predicting which group would exceed in a similar repeated study, we are unable to rely on the difference as being absolute. It may be true that a group of people such as these inactive teachers would always surpass in Tests and Measurements a group such as these active teachers under similar conditions but we are unable to make such a statement due to a low critical ratio found as a result of too few people represented in the study. As pointed out before, however, the consistency of results, even in the absence of a sufficiently high critical ratio, gives some indication of possible trends in future studies.

A summary and conclusions will appear in Chapter VI.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

In this chapter are summarized the individual conclusions indicated by the data in Chapters IV and V. Recommendations will be made on the basis of these conclusions.

Part 1. Test Analysis:

1. The test was found to be a reliable measure of the knowledge and abilities which it measures. For the three sittings the reliabilities were respectively: .95, .90 and .92, and .78. Probable errors ranged from .01 to .05. The First and Second Sittings are very reliable in their present form. The Third Sitting should be increased in length to increase reliability.

2. The first two sittings are good measures (.96 $\pm$ .01 and .94 $\pm$ .02) of the criterion, Total Score. The Third Sitting because of its short length and scoring procedure is not a good measure of the criterion in its present state, but since it has value as a device for measuring different abilities than those measured by course grades and those measured by the first two sittings, a recommendation will be made to increase its length and change its scoring. This procedure should make the sitting a better measure of the criterion.

3. On the standardized Second Sitting, the best measures of the criterion, Total Score, are Educational Psychology (.82 $\pm$ .04) and Principles of Teaching (.81 $\pm$ .04). The other measures are Tests and Measurements (.68 $\pm$ .07) and Principles of Organization and Management (.65 $\pm$ .07).

4. The First and Second Sittings, which contain material from some of the same courses, measure to a rather high degree (.84 ± .04) the same knowledge or abilities. This fact raises the question as to whether it is advisable or necessary to retain both sections.

5. The Third Sitting does not measure to any great extent the same knowledge or abilities as the first two sittings. Correlations of the Third Sitting with each of the first two sittings were .54 ± .09. Although the Third Sitting does not correlate highly with the criterion, Total Score, it is found to have value, as mentioned in Part 2. Possible reasons why the sitting did not correlate highly with the criterion are: a) its comparatively short length which results in limited sampling; b) its short range of possible scores; c) the criterion was composed principally of other types of measurement.

6. On the First Sitting the True-False and Multiple Choice sections correlate highly (.97 ± .01 and .76 ± .05) with the criterion, and do not measure the same abilities (inter-correlation .64 ± .07). It is therefore advisable to retain both.

7. On the First Sitting the completion and matching sections do not correlate highly with the criterion (.56 ± .09 and .64 ± .08). It is therefore advisable that they be discarded.

Part 2. Examination Analysis:

8. On the standardized Second Sitting, the graduate students are distributed from a mere passing grade on under-

graduate subject matter to a level which exceeds 94% of the students in undergraduate education courses throughout the nation.

9. The average graduate student who took the Examination did not score as highly on the sections of the standardized sitting as the average undergraduate student throughout the nation. The percentiles attained in the various sections were: Educational Psychology 13, Tests and Measurements 39, Principles of Organization and Management 41, Principles of Teaching 33, Total 24.

10. The answer to the question, "What is the relationship between scores on the Comprehensive Examination and success in college courses in Education?" when the average grades received by students varied from 72 to 92, is as follows.

The relationship between Average Grades and Examination sections are: Total Score .67 $\pm$ .07, First Sitting .62 $\pm$ .08, Second Sitting .69 $\pm$ .07, Third Sitting .43 $\pm$ .10. The Examination does not measure essentially the same thing as do grades in college courses. It would therefore appear worthwhile to retain the Examination as a supplementary device for measuring the students' grasp of the general field. The sitting which bears the least resemblance is the Third Sitting which is composed of case problems. The Third Sitting, in which the student is forced to think in terms of educational principles, therefore presents the best means for measuring something different from what is measured by scholastic grades.

The relationship of the number of Education courses taken and the Total Score was .21 $\pm$ .12 which shows that with an

increase in the number of graduate courses taken, there is practically no probability of an increased score on the Examination.

11. The answer to the question, "What is the relationship between scores on the Examination and the number of years since graduation from undergraduate college?" where the number of years varied from 1 to 28 with a concentration at 1 to 6 years, is the fact that there were no significant relationships between number of years out of college and any section of the Examination. Correlations ranged from -.12 $\pm$ .12 to .36 $\pm$ .11. However, the controlling element in these correlations was the large group of students who had been out of college less than 6 years and who, because of their great variability, overshadowed the few students who had been out of undergraduate college from 6 to 28 years. This distribution of years out of college and the fact that graduate students here did not do as well as the nation's undergraduates would appear to call for further study regarding the relationship of success on the Examination and the number of years out of college. The relationship between the number of years out of college and Total Score, when found for active teachers alone, was .05 $\pm$ .16, which substantiates the results found for all the students and indicates that the amount of time since graduation from undergraduate college is not a major factor in Examination success.

12. The answer to the question: "What is the relationship between scores on the Examination and the type of undergraduate college attended?" follows. Since there were too

few subjects on the basis of which to predict what other graduates of various type colleges would do, the grouping was regarded as a grouping of individual graduates of these types of colleges. The outstanding evidence was the consistency of results on the Examination, which results agreed essentially, in rank, with college grades. The conclusion was drawn that the ruling factor was general ability of the student. The only outstanding exception in all work done on this question was the high relative ability of the Teacher's College graduates on the Third Sitting case problems. This ability was credited either to special training or special ability along that line. General ability still held control for the other four groups of students. Most of the evidence placed the institutions in this order: 1) Engineering, 2) Liberal Arts, 3) Art, 4) Physical Education, 5) Teachers' Colleges.

13. The answer to the question, "What is the relationship between scores on the Comprehensive Examination and in-service versus prospective teachers?" is as follows. The active teachers surpassed the prospective teachers in all sections except one, Tests and Measurements, which a majority of active teachers had not studied. Again the results are consistent, and superiority on the Examination is in the same direction as superiority in grades received in college courses. Since correlations of scores with the number of years out of college were not found to be significant for these people, two possibilities exist to explain the superiority of the active teachers: either the active teachers were superior

individuals or they had in teaching gained something which the Examination measures. Realizing that for this limited number of people differences are not reliable, and using consistency of results as a possible indication of effects, we might say that since superiority in grades is in the same direction as superiority in scores, there is no indication that the active teachers gained a measurable quality in teaching.

14. The answer to the question: "What is the relationship between scores on the Examination and imminence of graduation in the year the Examination was taken?" follows. Although the students who received the Master's degree had a slightly higher average of college grades (83.48 as against 82.01) than the students who did not complete the requirements for the degree, they were lower than the Bachelors on all sections of the examination except Principles of Organization and Management.

15. The answer to the question: "What is the relationship between scores on the Comprehensive Examination and sex of the students taking the Examination?" is as follows. The women led the men on all sections of the Examination except on the section on Educational Psychology on which the two groups were equal to the second decimal place. In college grades the men were 0.37 point higher. This slight variation is in the opposite direction from the variation in Examination scores. This difference makes the second slight exception to the rather remarkable consistency of results. The only possibility of explanation of this slight exception is the

lack of reliability of differences. The men who took the Examination were more variable in that they scored higher and lower than the women.

16. The answer to the question: "Are the high ranking students superior on all parts of the Examination?" is that the 5 highest ranking students are very noticeably superior to the 5 lowest ranking students on all sections of the Examination and on average grades received in college courses. The reliability of the differences was not calculated on the basis of groups of 5 people each, but again the consistency of results points to general ability as the controlling factor even though ability as measured by grades may not give exactly the same rank as ability measured by Examination scores.

17. The answer to the question: "Are differences between groups significant?" is as follows. The question, when reworded appears thus: If the study were repeated with the same materials and with similar subjects, would the same groups again exceed the groups which were found to be relatively poorer in the first study, or, in the second study, would the situation be reversed such that the groups formerly considered poorer would now exceed? Due to the fact that the number of people involved in this study is limited to 30, the differences obtained in this study are not reliable; i.e. with an increased number of students the same groups might be superior, but we are not able to state the fact definitely as a result of this study. As pointed out, however, the consistency of results, even in the absence of a sufficiently high critical ratio, gives some indication of possible trends in future studies.

Recommendations: The following recommendations are a personal evaluation of the foregoing conclusions in an attempt to revise the Comprehensive Examination as a result of this study.

1. Since the Examination does not measure essentially the same thing as do course grades, that it be retained as a supplementary device for measuring the student's grasp of the general field of Education.

2. Since the Completion and Matching sections of the First Sitting do not have a high relationship with the criterion, Total Score, that they be discarded.

3. Since, exclusive of the discarded sections, 178 items on the first section were found to be statistically poor for reason of being too easy, too difficult, or to be answered better by the poorer students, that these items be inspected by the Department personnel for an evaluation as to validity and proper wording.

4. Since the First and Second Sittings measure to a rather high degree the same knowledge or abilities, that in the revision of the First Sitting the Department personnel attempt to substitute such new items that the sitting may tend to measure to a greater degree different knowledge or abilities. This revision might take the form of a greater proportion of graduate material arranged in groups to allow the student a choice according to the particular courses he has pursued in his graduate work.

5. Since the Third Sitting was found to measure different knowledge or abilities than the other two sittings, and since the Third Sitting was found to be the best means

of measuring something different from what is measured by course grades, it is recommended that the sitting be retained.

6. Since the Third Sitting had so few items its reliability was comparatively low, and it had a relatively small effect upon the Total Score. The time necessary to take this sitting was one-third or one-quarter the time necessary for either of the other two sittings. It is recommended that the number of case studies, of a similar type, be increased to between 70 and 90, and that scores obtained on the revised sittings then be multiplied by a constant, probably 2, in order to give a range of scores approximately equivalent to those obtained in each of the other two sittings.

7. Because of the relatively weak grasp of the general field of education by graduate students as a group, it is recommended that one form of the standardized Second Sitting, or a similar standardized test that covers the basic courses in the field of Education, be administered to each student upon entrance to the Graduate School. This test is not to be an admission requirement but to ascertain whether the student has a sufficiently broad grasp of the undergraduate courses common to the subject. If the student does not on each section reach some arbitrary standard such as the average attained by undergraduates whose scores are in the published norms, it is recommended that the Department specify that he take the corresponding courses in this college either with graduate credit if such credit is given in those courses, or without credit if no graduate credit is given in the particular course or if the student has previously received credit in such a course.

8. In agreement with Boyd (20) who writes of the aims of comprehensive examinations at the University of Kentucky it is believed that "the student will be motivated to begin mastery of a field rather than to pass a succession of disconnected courses". It is believed that the Comprehensive Examination should be a criterion for the degree: a) in order to insure such motivation and increase student morale; b) because it measures something different than course grades; c) in order to exclude, at least temporarily, the student who is far below the general run in his grasp of the field. It is recommended that the passing mark on the Examination be determined separately for each revision in this manner:

A. That the Total Scores on the Examination be arranged in a frequency distribution to contain between 10 and 15 classes.

B. That the passing mark for the particular revision of the Examination be placed at the lower limit of the class which marks the end of the regular decrease of frequencies, provided that that lower limit is at least as far below the class which contains the average of the group as the lower limit of the class containing the highest Total Score is above the average of the group.

9. It is recommended that any student failing the Examination be considered by the Department personnel for eligibility for the degree on the basis of previous attainments and work. If previous attainments were low and work was poor, it is recommended that the student be advised as to his greatest weaknesses and that he be given a second opportunity to take the Examination a half year or year later. A passing score on the second trial should satisfy requirements, and failure on the second trial should mean rejection.

APPENDICES

APPENDIX I

THE SPEARMAN-BROWN FORMULA

The Spearman-Brown formula (27) expresses the relationship between the length of a test and its reliability. The general form is:

$$r_n = \frac{n\, r_{xy}}{1 + (n-1) r_{xy}}$$

In the formula r_n is the reliability that may be obtained by n applications of the test, and r_{xy} is the coefficient of correlation for the first two applications actually made. Thus, if a test were applied twice, the reliability for any number of applications could be calculated.

So far as reliability is concerned, there is no theoretical difference between giving a test, say, ten times or giving a test which is ten times as long as the given test. If the correlation between two applications of a test was .75, the reliability obtained by 10 applications, or a test 10 times as long, would be:

$$\frac{10 \times .75}{1 + (10-1)\ .75} = .97$$

In this thesis reliabilities are found by the Odd-Even method which considers that the odd items may be taken as one test, and the even items as another test. Since each group composes half the test, their correlation will be the relia-

---

27. Odell, Charles W. Statistical Method in Education, New York: D. Appleton Century. 1935. pp 209-14.

bility for half the test. It will be necessary to correct that correlation by the Spearman-Brown formula to find the reliability of the whole test.

The reliability for half Form I of the standardized Second Sitting was found by odd-even correlation to be .82. By the Spearman-Brown formula, the reliability for the whole form is:

$$\frac{2 \times .82}{1 + .82} = .90$$

Each of the other odd-even correlations was corrected by the formula to give the reliability of each whole sitting.

APPENDIX II

THE INDEX OF RELIABILITY

When the odd items of a test are considered as one test and the even items are considered as another test there are likely to be errors in both sets of measurements. In a single set of measurements there is only one set of errors. The correlation between one set of measurements and the theoretically true measurements is lowered less by one set of errors than the correlation between two sets of measurements is lowered by two sets of errors. The coefficient of correlation between the theoretically true scores and the obtained scores is called the "Index of Reliability" (28) and is calculated by extracting the square root of the coefficient of correlation between the two sets of scores. The Index of Reliability is always greater than the coefficient of reliability unless both equal 1.00 or zero.

In this thesis the Index of Reliability for each sitting was obtained by taking the square root of the reliability obtained for each sitting by the Spearman-Brown formula. (The effects of the two sets of errors in the odd and even groupings are present, of course, in this reliability.) For example, the reliability for Form 1 of the standardized Second Sitting was found to be .90. The Index of Reliability is the square root of .90, or .95.

---

28. Ibid: p.215.

APPENDIX III

THE SIGNIFICANCE OF CORRELATIONS

Having found the relationships between sittings and sections of the test, and between scores and other factors which might be related to scores, it is of interest to try to evaluate those relationships.

Of the two possible units for the measurement of reliability, "Standard Error" and "Probable Error", the Probable Error, PE, was chosen as the unit in this study because of its universal use. If the PE be considered with relation to the normal curve, it is that horizontal distance from the mean which, when taken in both a positive and negative direction from the mean, projects half the area under the normal curve. Therefore the chances are even, 1 to 1, that any particular score will be within plus or minus one PE from the mean. When a distance of 4 PE from the mean is taken, the chances are about 997 out of 1000 that any score would lie within those limits.

With regard to correlation coefficients, we say that any coefficient which is greater than 4 PE is statistically "significant", in other words there is some relationship between the two variables not "entirely accounted for by chance fluctuations in random sampling" (29).

---

29. Lindquist, E.F. A First Course in Statistics. Boston: Houghton Mifflin Company. 1938. pp xi + 226 (p177).

However, statistical significance is not necessarily true significance for the field of study or the particular study in question. Thus, a coefficient of .90, while practically unattainable in some fields, would be considered too low to be usable in regard to the reliability of an individual intelligence test. Certain fields of study require a certain magnitude of the coefficient to assure significance in these particular fields. For this study, a rough table of significant values follows. In each case, statistical significance is necessary before investigating the valuable significance.

Table XXXI

Correlation Coefficients of Valuable Significance in Cases Where Statistical Significance Has Been Obtained

.60 - .80	Significant
.80 - .90	Highly Significant
.90 - 1.00	Very Highly Significant

BIBLIOGRAPHY

Books, Theses, and Tests

Bracy, A. A. — A Study of the General Achievement of Prospective Teachers at Massachusetts State College. Unpublished Master's Thesis, 1939.

Chamberlain, L. M. — The Teacher and School Organization. New York: Prentice Hall Inc. 1936. pp. xxviii ± 656. Chapter 5.

Jones, E. S. — Comprehensive Examinations in American Colleges. New York: The Macmillan Company. 1933. pp xix ± 436.

Lindquist, E. F. — A First Course in Statistics. Boston: Houghton Mifflin Company. 1938. pp. xi ± 226.

Newson, N. W., Smeltzer, C. H. and Bowden, H. T. — Comprehensive Examination in Secondary Education. Keokuk, Iowa: The Extra-Curricular Publishing Co. 1938. Forms 1 and 2.

Odell, C. W. — Statistical Method in Education. New York: D. Appleton-Century. 1935. pp xx ± 457.

Tiegs, E. W., and Crawford, C. C. — Statistics for Teachers. Boston: Houghton Mifflin Company. 1930. pp xvi ± 212.

Ullman, R. R. — The Prognostic Value of Certain Factors Related to Teaching Success. Ashland, Ohio: The A. L. Garber Co. 1931. 133 p.

Periodicals

Bent, R. K., and Douglass, H. R. — Differences in the Performance of Departmental Groups of Student Teachers on Qualifying Examinations at the University of Minnesota. tabs. School and Society, 45:726-7, May 22, 1937.

Birkoff, G. D. — General Final Examinations at Harvard College. School and Society, 43:86, January 18, 1936.

Boardman, C. W., and Patterson, D. — Achievement Examinations as a Technique for Selecting Students in Education. tabs. School and Society, 49: 586-8, May 6, 1939.

Bossing, N. L. — Teacher Aptitude Tests and Teacher Selection. Research in Higher Education. Washington: U.S. Dept. of Interior, Office of Education, General Printing Office. 1932. pp 117-33.

Boyd, P. P. — Compulsory Comprehensive Examinations. School and Society, 37:817-20, June 24, 1933.

Eldred, Lewis — Traits of a Group of Prospective Teachers in the College of Arts and Sciences, Cornell University. School and Society, 50:477-50, October 7, 1939.

Hardesty, C. F. — Can Teaching Success be Rated? Nation's Schools, 15:27-8, January, 1940.

Kriner, H. L. — Second Report on a Five Year Study of Teachers' College Admissions. Educational Administration and Supervision, 21:56-60, January, 1935.

Lockwing, W. F. — Individual Plan of Graduate Study at the University of Iowa. School and Society, 30:743-5, November 30, 1929.

Major, C. L. — Percentile Ranking on the Ohio State University Psychological Test as a Factor in Forecasting the Success of Teachers in Training. School and Society, 47: 582-4, April 30, 1938.

Mathews, L. H. — An Item Analysis of Measures of Teaching Ability. Journal of Educational Research, 33:576-81, April, 1940.

Moore, C. H. — Requirements for Graduate Study at Harvard University. School and Society, 29:575-6, May 4, 1929.

Phillips, G. B. — Teacher Training Deviates at North Carolina. Nation's Schools, 22:29, December 6, 1938.

Pilgram, R. J. — The Comprehensive Final Examination Plan Adopted for Franklin and Marshall College Seniors. School and Society, 33:60, January 10, 1931.

Sandiford, P., et al. — Forecasting Teaching Ability. Elementary School Journal, 38:326-9, January, 1938.

Steiner, R. W. and Van Haden, H. I. — The Pretraining Selection and Guidance of Teachers. Journal of Educational Research, 33:321-50, January, 1940.

Stumpf, W. A. — Graduate Work in State Teachers' Colleges. tabs. School and Society, 46: 834-8, December 25, 1937.

Townsend, M. E. — An Experiment in the Professional Examination of Teachers. School and Society, 50:537-41, October 21, 1939.

Troyer, M. E. — The Selection of Students for the Profession of Teaching. Journal of Educational Research, 33:581-94, April, 1940.

Whitney, F. L. — The Prediction of Teaching Success. Journal of Educational Research. Monograph #6. Bloomington, Illinois: Public School Publishing Co. 1924. 85 p.

——— — Qualifying Examinations in the College of Education at the University of Minnesota. School and Society, 35: 698, May 21, 1932.

ACKNOWLEDGMENTS

The writer wishes to express his sincere gratitude and appreciation to the thesis committee: Dr. A. W. Purvis, head of the committee, Dr. A. E.Andersen, and Dr. W. M. Miller, for their active interest, co-operation, and constructive criticism.

Approved by:

Allen P. Anderson

Walter M. Miller

Albert W. Purves

Thesis Committee

Date June, 1940